图书情报工作
第 64 卷 第 4 期 2020 年 2 月

科技文献知识基因表达及遗传与变异研究*

■ 白如江[1,2] 张庆芝[3] 孙一钢[1]
[1] 国家图书馆 北京 100081 [2] 山东理工大学科技信息研究所 淄博 255049
[3] 北京大学信息管理系 北京 100871

摘 要：[目的/意义]知识的积累与传承推动着人类社会的发展,为此提出利用科技文献知识基因进行知识的遗传与变异研究,以期对知识传承与发展变化进行更直观、全面地透视。[方法/过程]通过辨析知识基因概念,确定知识基因的研究意义,并探讨其具体研究对象;提出科技文献知识基因内容表达的两种方式,归纳科技文献知识基因的类型;分析影响知识基因遗传与变异的主要因素,且设计识别知识基因的遗传与变异的方法。[结果/结论]通过对科技文献知识基因的辨识,能够有效揭示出不同文献之间的知识传承与迭代,促进知识基因理论体系的发展与完善。

关键词：模因 知识基因 知识基因遗传 知识基因变异 文本内容分析
分类号：G250
DOI：10.13266/j.issn.0252-3116.2020.04.009

"原子(atoms)"、"字节(bytes)"及"基因(genes)"被认定为 20 世纪最具颠覆性的科学概念[1]。而 3 项概念在结构上皆为最基本的组织单元:原子是物质的最小单元,字节(比特)是数字信息的最小单元,基因是生物信息与遗传的最小单元。为什么这些最小单元充满了独特的魅力? 根据分形理论(fractal theory),理解最小单元的组成可把握整体情况[2]。人们只有充分理解基础概念,才能领悟其特性或演化规律。字节与基因只是人类发明的符号,精妙的自然法则是数字化信息理论的支撑[3]。基于字节与基因共同遵循的自然法则规律,可相互支撑对相关领域的科学知识,产生新的领域。

基因的研究基于碳基生命,知识基因的研究则是基于人类科研成果最重要的载体——科技文献。本文通过对知识基因狭义及广义定义的梳理,明确知识基因的研究意义,并探讨其具体研究对象;分析知识基因的类型、影响知识基因遗传与变异的因素,展望知识基因的未来发展方向,促进知识基因理论体系的发展与完善。

1 知识基因定义

道金斯最早在基因与知识之间建立联系[4]。他认为存在脱离化学物质的"基因",这种"基因"的"汤"为人类文化,传递载体为语言,产卵场为大脑。他赋予其新的名字"模因(meme)"。模因产生于大脑,并在纸张、胶卷、硅片等信息载体上建立滩头堡。简单的想法如颜色及数字等类似于单个核苷酸,不足以成为模因。模因应像基因一样是复杂的单元,并具有一定的持久力[5]。在道金斯提出 meme 概念后,meme 在国内外得到广泛关注与研究。

在国外,早期遗传学家利用基因的遗传与变异得出阶级分层不利于推进社会公平的结论[6],因而,大部分社会学家对基因学说持抵触态度。此后,meme 被用于文化研究。S. K. Sen 在文化基因基础上提出情报基因的概念,情报通过情报基因遗传、试错检验等逐渐稳定与增长[7]。S. Blackmore 在 The Meme Machine 一书中定义知识基因是高保真复制,高繁殖力和生命力强的复制子[8]。R. Aunger 在 The Electric Meme 书中从语言学角度分析 meme,认为流行语也是一种 meme,强调 meme 在流行文化中的解释作用[9]。在这些研究中,学者未划分其研究边界、缺乏严格定义、难以将研究数学化,从而使文化基因过于抽象、无形、不可度量[10],只能在文化进化理论及心灵进化理论的研究中起解释作

* 本文系中国博士后科学基金项目"基于知识基因表达的科技创新路径识别研究"(项目编号:2018M640101)研究成果之一。
作者简介:白如江(ORCID:0000-0003-3822-8484),副研究馆员,硕士生导师,博士;张庆芝(ORCID:0000-0001-9161-9754),博士研究生,通讯作者,E-mail:zhangqingzhi@pku.edu.cn;孙一钢(ORCID:0000-0001-8478-1737),副馆长,研究馆员,博士。
收稿日期:2019-05-08 修回日期:2019-10-08 本文起止页码:78-87 本文责任编辑:徐健

用,而无法开展应用实践。

在国内,李伯文认为知识基因就是科学概念[11]。刘植惠认为此定义过于宽泛,从而对知识基因重新定义:知识基因是知识进化的最小功能单元,具有稳定性、遗传与变异性、统摄性、指向性,其目的是摸清知识进化规律[12,13]。刘植惠详细阐述了知识基因的定义、特征、分类、科学定律、应用及遗传运动与变异运动等[14]。孙晓玲将在文献中经常出现的词语或短语定义为重要的知识基因,并结合文献中词语出现次数及在引文网络中的传播程度,计算知识基因强度[15]。刘则渊认为知识基因是在特定知识领域所构成的自组织知识系统,可展示出知识的产生、演化与重组、涌现、断层和变革、传播和应用等[16]。和金生等人研究了知识基因在企业创新过程中的作用与反馈机制[17]。顾新建等使用 SAO 三元组方法从专利引证网络中提取知识基因并建立知识进化轨迹[18]。丁堃等使用知识基因发现算法识别知识进化与冲突中起关键作用的知识基因[19]。谭宗颖等构建了基于知识基因游离与重组的主题演化研究模型,以了解学科领域的发展和演化规律、学科领域的研究主题布局[20]。

由上述知识基因发展历程及应用场景可知,知识基因是知识进化轨迹的基本单位,可利用知识基因发现和挖掘隐含的、未知的、潜在的有价值知识,为知识创新提供智力支持。但是在知识基因概念描述方面较为模糊,仅使用解释性语言描述知识基因理论及知识基因识别算法能够完成的任务,但并未说明知识基因是什么,为什么可完成任务等重要知识。虽然文章标明知识基因,但在实际应用过程中仍使用主题词、关键词等表征知识基因。

笔者认为,科技文献知识基因是模因的意义表达类型之一,是科技文献文本内容中表征文献价值的知识对象的有机结合体,是科技创新中最基本、最活跃、影响面最宽的知识内容。知识基因由原始文献摘要及施引文献引文内容组成。由于施引文献在对原始文献引用时,引用内容并不一定是原始文献最突出的创新点,通过二者的有机结合、重组可以得到科技文献的知识基因。

科技文献的知识基因生成过程应该由以下几个步骤:首先,提取原始文献摘要及施引文献引文内容的类别标签;其次,识别引文内容的引用位置、引用情感、引用功能、引用性质;最后,根据上述标注计算遗传与变异结果,见图 1。

在上述知识基因生成基础上,通过综合引用类别

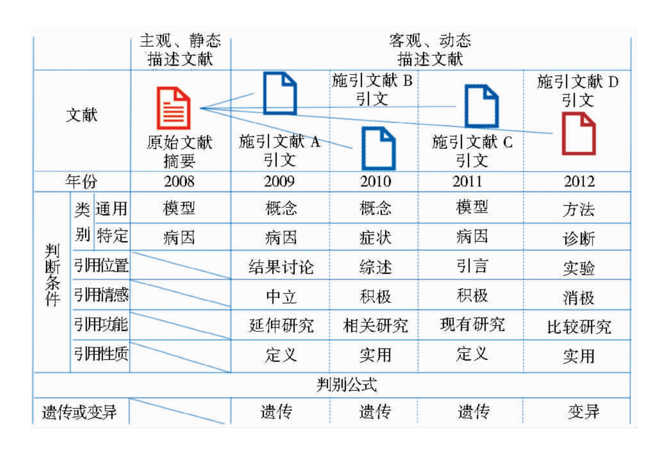


图 1　科技文献知识基因生成过程

标注结果、遗传与变异判断结果、时间等因素完成单篇科技文献知识基因的表达提取,如图 2 所示。图中横坐标为特定知识基因类别标签,纵坐标为通用知识基因类别标签,黄色表示原始文献,红色表示施引文献的引文内容与原始文献为知识基因变异关系,蓝色表示施引文献的引文内容与原始文献为知识基因遗传关系。在黄色、红色或蓝色方块内数字为科技文献的发表年份。

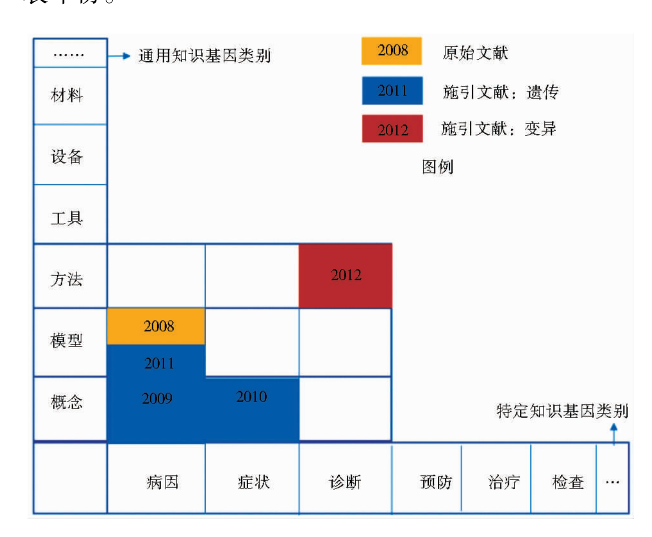


图 2　科技文献知识基因表达形式

2　知识基因与知识元

吴军在《见识》一书中提到,伟大的发明总是基于前人的工作,它所完成的是从“N”到“N＋1”的过程[21]。基于现代技术的飞速发展,实现从 0 到 N 的全过程变得越发不可实现。当前科学发展模式为在原来的基础上更进一步,站在巨人的肩膀上(N),产生新的技术或工具(1),实现质的飞跃,靠着一点点的进步,推动科技发展。同样,人类在习得知识过程中并非将前人所做设想与实验一一验证,而是直接将部分知识

设为定理等,使人类快速站在当前技术发展前端,致力于新知识的产生。在科技文献中,N 可理解为原始论文的被引文献,1 可理解为原始论文的摘要。当论文被施引文献引用时,原始论文成为"N"。科技文献在"N"到"N+1"的过程中不断演变与发展,因此单篇文献的知识基因应由原始文献的摘要及其施引文献的引文内容组成。

由上述可知,知识基因是动态的、发展的,而不是一成不变的。而目前通常讲的知识元、知识单元等概念都是指具体实际的词语,这些词语或是主题词,或是三元组。知识元通常是通过对科技文献内容利用信息抽取、主题模型等技术抽取出的固定存在。在对知识元的研究中,许多学者聚焦在扩充主题词的属性描述与关系描述,如将文献的来源、版本、评论、访问记录等信息用于表示知识元,其主要目的是用于表示知识,控制与处理知识,其最终仍是碎片化知识[22]。

知识基因在表示知识时,使用通用知识基因类型标签与特定知识基因类型标签,以防止因知识的碎片化而迷失在知识的海洋中。同时,特定知识基因类型标签根据知识领域的不同而不同,从而使其较为自由灵活而不被限制。两种标签除了描述原始文献知识基因外,还用于描述施引文献引文内容。由于引文内容的可计算特点,可判断知识在由原始文献向施引文献发展的过程中,发生了遗传还是变异。因此,知识基因在表示原始文献的知识同时,引入了施引文献的内容与关系,内容使用两种标签类型表示,关系使用遗传或变异表示,随着施引文献不断产生,原始文献的知识基因的内容不断丰富。

3 知识基因的组成

人体的体细胞内有 23 对染色体,但对性别具有决定作用的染色体只有一对,即第二十三对染色体[23]。当科研人员的研究重点为性别等信息时,科研人员会着重研究性染色体。相似的,在知识领域,单篇文献会涉及到多个知识点,如对解决问题的各种方法综述,各个研究方法的实现工具综述等,但文献的突出贡献点是在前人研究基础上在方法等方面的创新。为了能准确描述单篇文献的突出贡献,知识基因的价值应由知识输出"N+1" 中的"1"决定。基于当前科研人员无法快速定位所需知识的普遍现象,结合基因与"N+1"法则,笔者认为,知识基因基于"N+1"法则中的创新点"1",是原始论文中表征原创观点、方法、数据、结论等有价值的知识单元的有机结合体。知识是流动的,笔者认为知识基因应由文献发表时作者对文献创新性的主观且静态的价值表达及施引文献中对文献的客观及动态的价值组成。静态表达体现在文献的摘要之中,在其发表时已经固定不变;动态客观描述体现在后续施引文献的引文内容之中,并客观描述施引作者所研究领域的客观价值。

3.1 知识基因的静态表达

论文在发表时,科技文献自身价值已经得到作者本人、编辑及同行评审专家的认可,构成其静态描述。作者在摘要之中将文献的主观价值进行精炼化表达,不加评论和补充解释,最大限度对单篇论文重要信息进行简短、扼要而连贯的陈述,集中反映原文精华。摘要具有简洁、明了、独立性、完整性、高度概括性和自含性等特征,并拥有与文献同等量的主要信息[24]。

笔者认为,摘要的本质是通过除去不重要的内容来压缩文献字数,保留文献目的、方法、结果、结论等信息性内容,表达科学、结构合理、逻辑严密,可独立存在成为二次文献。摘要是作者视角下,单篇文献的价值表征。通过对摘要细粒度化、结构化、语义化表达,可准确描述知识及其相互之间的关系。因此,知识基因的静态表达应从文献的摘要之中提取。

3.2 知识基因的动态表达

施引文献作者通过引用对被引文献客观价值进行升华,随着后续其他科研人员的引用与评价构成文献后续的动态价值。科研人员的研究方向及看待文章的视角决定了这篇文献对施引作者的贡献,构成文献后续的客观描述。

引文内容是施引文献基于当前发展状况,对被引文献知识创新点的客观判断,反应科学知识的递增性规律,在表征被引文献对后续研究的主要贡献同时,相比于被引文献摘要和全文,能够提供更加客观和丰富的语义信息,表达施引作者对原始文献价值的认识[25]。因此,笔者认为引文内容作为科学传播与交流的痕迹,知识基因的动态表达需要从引文内容中提取。静态摘要与动态引文内容的功能表现如表 1 所示:

表 1 摘要与引文内容的功能分析

功能表现	摘要	引文内容
视角	作者视角	读者视角
描述角度	主观价值	客观价值
表现形式	静态表现	动态表现
表现层次	宏观层次	微观层次
侧重点	创新性	知识基础
分析维度	单一维度	5 个维度
分析依据	内生指标	外生指标

白如江，张庆芝，孙一钢. 科技文献知识基因表达及遗传与变异研究. [J]. 图书情报工作,2020,64(4):78－87.

4 知识基因类别

知识与知识之间的连接构成一张复杂的知识网络[26]，在知识网络中对某一知识点的定位需要横坐标与纵坐标的结合。本文提出利用通用知识基因作为纵坐标，特定知识基因作为横坐标进行知识基因可视化表达。

通用知识基因是各研究领域科技文献通用的知识基因对象，比如研究方法、研究目的、理论、工具、数据等。

特定知识基因是基于研究领域特点的知识基因对象。以 Alzheimer 研究领域为例，特定知识基因对象包括疾病症状、检测方法、治疗方案等。

作为基因的载体，染色体在不同种类生物中数量不同且恒定，正如各个学科领域的研究对象数量固定且不同。因此特定知识基因对象的设计需要一定专业知识，根据领域特点选取研究对象，或是在对研究领域有一定了解后，根据研究领域的知识库进行研究对象设计。叙词表及本体作为对某个领域知识的共同理解，将特定领域的实体概念及相互关系、领域的特性和规律进行形式化描述，在地球科学领域[27]、能源交通领域[28]、地质学领域[29]、气象领域[30]、生物医学领域内[31]广泛应用，可为特定知识基因研究对象设计提供参考。

4.1 通用知识基因类别

M. J. Moravcsik[32]、M. Garzone[33]、I. Spiegel-Rösing[34]、C. Oppenheim 等[35]、R. Radoulov[36]、陆伟[37]、秦春秀[38]等对科技文本内容通用标注对象进行了研究，见表 2。这些标注对象可以表征科技文献的通用知识基因。

表 2 科技文献内容标准对象

对象 \ 人员	M. J. Moravcsik	M. Garzone	I. Spiegel-Rösing	C. Oppenheim and S. P. Renn	R. Radoulov	陆伟	秦春秀
概念	√		√		√	√	
定义			√				√
解释、内涵			√				√
理论、原理				√	√	√	√
问题			√				√
数据			√	√	√	√	
材料		√	√				
设备		√					
条件		√					
工具	√	√				√	
方法		√	√	√	√	√	√
步骤、方案		√					√
假设			√				
算法						√	√
公式						√	
方程		√		√			
模型						√	√
系统							√
应用				√	√	√	√
结果		√				√	
未提及						√	

这些通用标注体系框架在实际应用过程中，如果区分过于详细，则对标注人员的区分能力要求过高；若是区分过于笼统，则失去通过标注实现细粒度知识组织的意义。因此，笔者结合前人研究，将区分度较低标注对象进行合并，选取以下标注对象作为通用知识基因表达对象：概念（包含解释、定义、内涵、原理、理论）、问题、数据、材料、设备、工具、方法（包含方案、步骤）、算法（包含方程、公式）、模型（包含系统）、应用、其他。

4.2 特定知识基因类别

不同的研究领域存在不同的特定领域知识基因，比如在医学领域，每一种疾病都有它的因果关系，当身体内的细胞、组织和器官发生能病理变化，或是生化反应出了问题时，就会反映在病人的症状和身体检查的

异常结果上。因此,身体内部的不正常变化,可以解释临床上所观察到的现象,在诊断疾病时,可利用外在的表现做线索,寻求致病原因[39]。Alzheimer 疾病的因果关系,至今仍为谜团,临床表现和病理变化表现的结合,是病理生理学探讨病因与结果的基础。笔者根据 https://medlineplus. gov/alzheimersdisease. html 及 https://www. nia. nih. gov/health/alzheimers 网站中对 Alzheimer 的基本知识分类介绍,结合电子病历的对象研究,设计了 Alzheimer 研究领域的特定研究对象。这些研究对象可以作为特定领域知识基因的类型表达,如表 3 所示:

表 3　Alzheimer 研究领域特定知识基因类型表达

知识基因类型	解释	具体表现
疾病症状	客观病态改变	记忆丧失、混淆时间地点、数值计算困难等
疾病病因	致病因子和条件	β-淀粉样蛋白、基因、糖尿病、抑郁症等
疾病检查	借助仪器等的化验分析	血检、神经影像检查、脑脊液检查等
疾病诊断	NINCDS-ADRDA 诊断标准	前期、中期、后期
疾病治疗	改善认知功能,控制进程	益智药、抗精神病药、促脑代谢药等
疾病预防	对健康影响的积极应对	行为矫正、生活能力培训、记忆能力训练等

5　知识基因的遗传与变异

J. MONOD. 认为,知识与生物体一样,通过融合、重组及分离维持其结构并繁衍生息。知识的"传播力"或"感染力"通过互动促进传播。淘汰也在演化过程中扮演重要角色,通过淘汰机制加速社会进步[40]。"传播力"或"感染力"在丹尼特眼中为"一辆在各个心智之间传递卓越知识的四轮马车"[41]。

知识基因的遗传表现为知识代际之间的传承,在惯性作用下形成固定价值,维持人类知识的稳定性。快速且动态发展的社会冲击着固化知识基因遗传,知识通过不断变异适应新环境,即知识的创新,知识的创新使原有知识不再按照原有知识发展路径进化,而是产生知识基因变异。知识基因的遗传与变异在交叉学科中表现最为明显。交叉学科是相邻学科间的理论交叉渗透、相互吸收、有机融合。各种知识相互碰撞产生知识的遗传与变异,使学术研究产生新动力,孕育重大科技成果。

5.1　知识基因遗传与变异影响因素分析

基因经过自然选择,产生遗传或变异现象,而知识基因在遗传或变异的过程中是人为选择产生的结果,基因通过子代的特性和性状表现对亲代的遗传或变异,而知识基因通过引文内容展示对原始文献的遗传或变异。这些人为选择在科技文献中体现为引文的基本性质,如:引用位置、引用情感、引用功能、引用性质等。

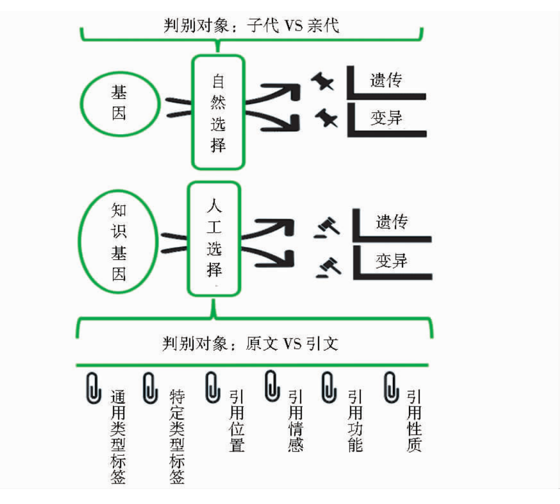


图 3　知识基因遗传与变异机制

科研人员根据自身的知识背景、研究方向、引用动机、引用目的、引用情感等对原始文献进行解读,正如"一千个读者就有一千个哈姆雷特",科研人员在进行引用时,受到自身研究主题、对被引文献的情感等主观因素的影响,使得文献在后续各篇施引文献中的价值不可等同化。在引文内容分析时,需考虑以下多种不同的因素:作者对被引文献的引用态度是积极还是消极;引文在文献中的位置可以表征引文的重要程度;引用的是文献中的概念还是文献的实际应用等。在每篇文献中这些影响因素的表现情况不尽相同,应将这些因素纳入考察范围,对引文内容进行分析建模计算。主要包括提取引用位置判断、引用情感计算、引用功能、引用性质、引用类别等。通过这些内容分析为知识基因的遗传与变异判别提供依据。引文内容对遗传与变异影响因素分析如表 4 所示,不同影响因素之间的关系见图 4。

下面具体分析不同影响因素。

(1)引用位置因素。H. Voos 认为将引用位置及引用功能相结合可计算引用的实际价值[42]。S. Teufel 经实证研究发现 62.7% 的参考文献对原文无实质性贡献,仅有 18.9% 的参考文献为原文献的术语定义或工具使用等部分内容提供价值输出[43]。B. A. Lipetz 对引文位置对文献的贡献进行实证研究,认为与介绍背景相关的引用对文献的科学贡献较小,而文献综述等可以指出当下问题存在哪里的引用对文献的科学贡

白如江，张庆芝，孙一钢. 科技文献知识基因表达及遗传与变异研究. [J]. 图书情报工作,2020,64(4):78-87.

表 4　引文内容对遗传与变异影响因素分析

一级影响因素	二级影响因素	影响因素描述
引用位置	引言部分 综述部分 方法论部分 实验部分 结果讨论部分	引文在文献中的位置不同,对施引文献的贡献价值也不同。综述部分引文数量多,主要是对发展历程的回顾,其重要性相对减少,知识基因遗传因素更多;而方法论等部分引文数量少,对问题的进一步探讨使其影响力增加,知识变异可能性增大。
引用情感	积极引用 中立引用 消极引用	引用情感体现作者对被引文献所做工作的正面、中立、负面的情感态度。正面态度遗传性大,负面态度变异性大。
引用功能	相关研究 比较研究 现有研究 延伸研究	相关研究及现有研究功能的引用多为简单提及,较为普遍,遗传性大;比较研究及延伸研究功能的引用可以激发新的想法,变异性大。
引用性质	定义类型引用 实用类型引用	引用性质体现论文的实质性输出内容形式,其贡献也不相同。定义类引用遗传可能性大。
引用类别	根据具体领域进行设计	见表 3 与前文

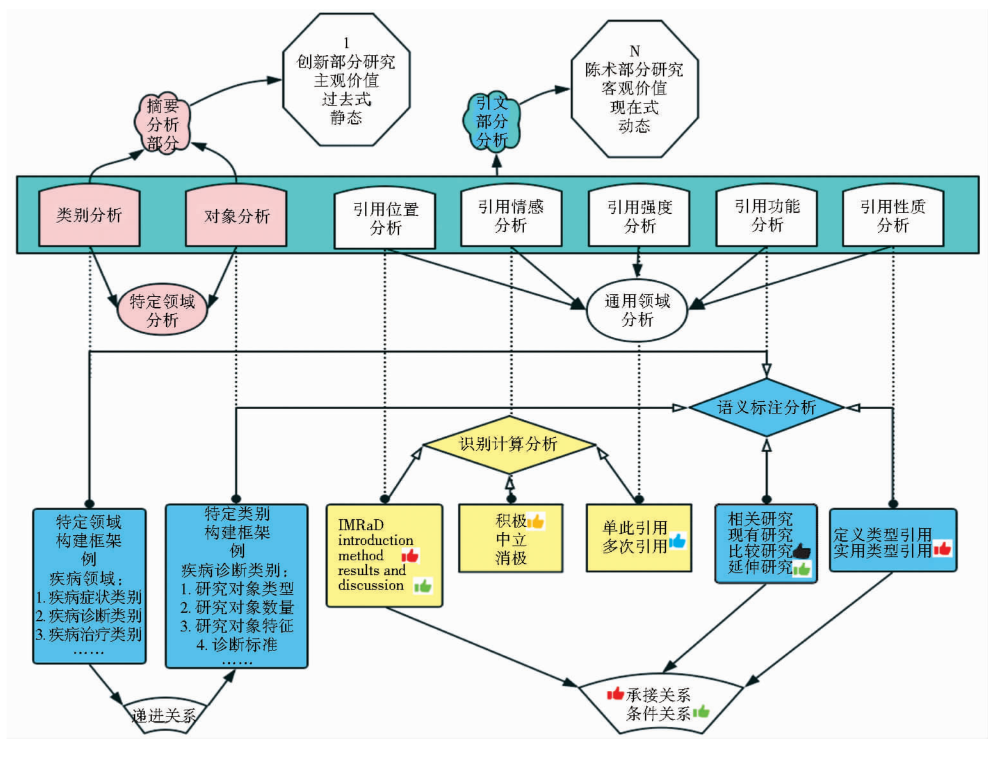


图 4　不同影响因素之间的关系

献较大,具有一定的引用意义[44]。G. Herlach 将引言与文献综述部分合并,将方法论与结果讨论部分进行合并,经过分析得到方法论与结果讨论部分对引证文献贡献更大[45]。X. Zhu 通过自动特征选择使用有监督机器学习方法建立了学术影响预测模型,该模型使用特征之一为位置特征,且这一特征在实验中发挥重要作用[46]。

通过上述研究分析,笔者认为出现在文章引言部分的引用对引文的科学贡献较小,在进行知识基因遗传与变异计算时需要降低变异计算权重,提高遗传计算权重。实验方法位置的引文则需要降低遗传计算权重,提高变异计算权重。

（2）引用强度因素。M. J. Moravcsik 对冗余引用进行实证研究,认为连续引用多为经典理论方法的引用,属于冗杂引用,作者对 30 篇文献 575 条参考文献进行分析后,发现 1/3 的引用为冗杂引用[32]。常思敏将连续引用判定为拼凑性冗余引用,并指出其对后文的叙述没有任何铺垫作用[47]。M. H. Macroberts 将引用分为有影响力引用与无影响力引用,研究结果显示有影响力引文数量较少[48]。

本文认为应将词语引用及连续度大于 3 的连续引用降低在遗传和变异计算过程中的权重。

（3）引用情感因素。D. E. Chubin 将引文分为积极引用与消极引用并进行人工判读，结果表明 95% 论文为积极引用[49]。M. J. Moravcsik 将引文分为积极引用、中立引用、消极引用，并通过设置问题进行人工界定，结果表明 84% 论文为积极引用[32]。在此之后，M. J. Moravcsik 将选择限定在积极引用与消极引用后，积极引用所占比例达 92%[32]。在国内，刘盛博将引用内容分为正面、中性、负面，对 BMC-bioinformatrics 期刊进行实证研究，结果表明 62.88% 的引用为中性引用，负面引用为 3.53%[50]。经上述研究表明，国外作者对积极引用的判断具有一致性，将大部分引用归为积极引用，与国内作者判断存在一定的差异。国内人员对积极引用的判断界定高，将大部分引用判断为中性引用。在对消极引用判断时，国内外作者在数量方面存在一致性，只有非常少数引用为消极引用。消极引用的文献虽然对原文进行引用，但原文所表达的主要思想并未得到传承，与原文献的原意相违背。

本文认为积极引用类型提高遗传计算权重，中立引用降低在遗传和变异计算过程中的权重，消极引用提高变异计算权重。

（4）引用功能因素。S. U. Hassan 根据前期阅读及句法结构特征，将引文功能划分为 4 种类型：相关研究、比较研究、现有研究、延伸研究。并汇总各类型提示词，使用正则表达式对引用进行分类，对其进行国家与机构的评价[51]。

笔者认为利用 S. U. Hassan 划分的相关研究、现有研究引用功能设计遗传计算权重，利用延伸研究、比较研究设计变异计算权重。

5.2 科技文献知识基因遗传与变异识别计算

根据前面辨析的知识基因类型和影响遗传变异的因素，本文设计了一种知识基因遗传与变异计算实现方法，如图 5 所示：

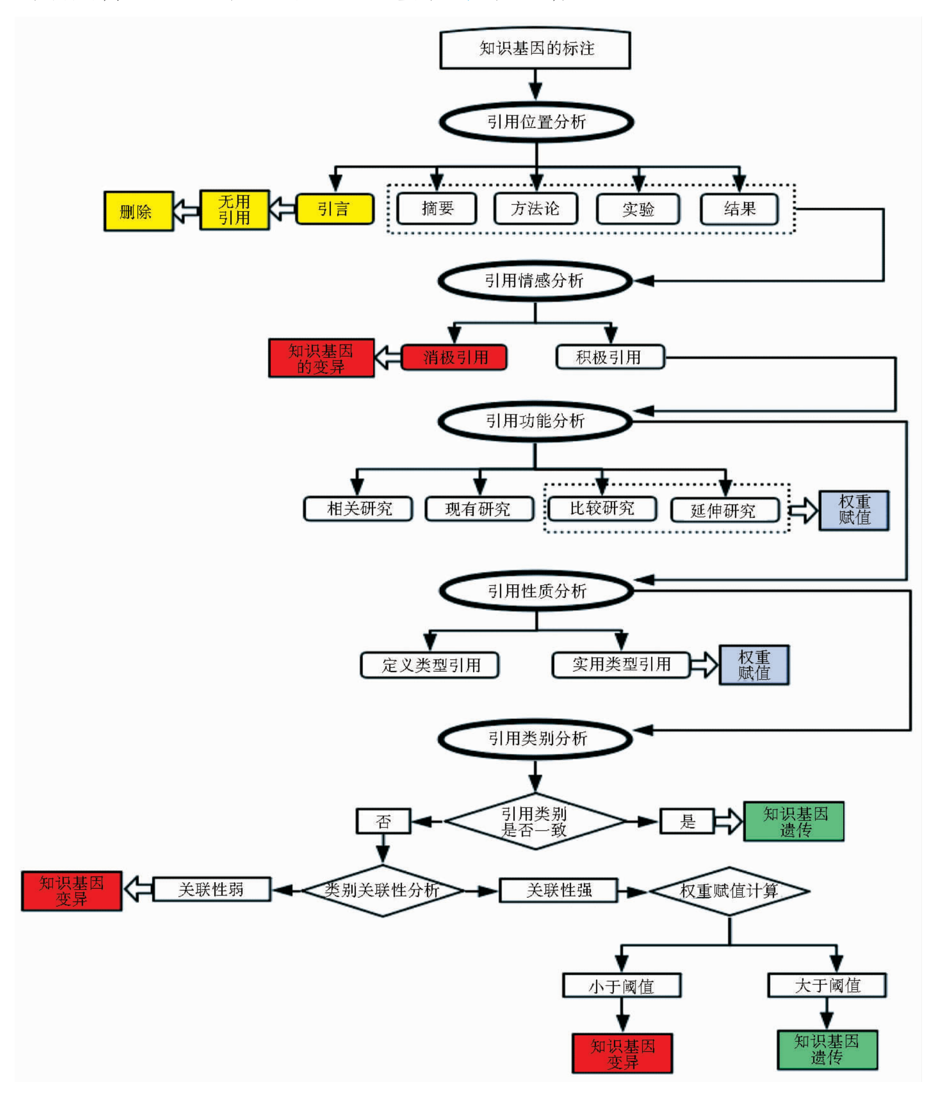


图 5 知识基因遗传与变异的计算研究

通过图 5 中计算步骤后，根据计算结果对知识基因进行遗传和变异综合判断，如果引用类别一致则判定为知识基因的遗传；如果不一致进行类别关联性分析，如果关联性弱则判断为知识基因的变异；如果关联性强，小于权重阈值的判断为知识基因的变异，大于权重阈值的判断为知识基因的遗传。权重与阈值的设置根据科技文献所属领域的文献特征决定。

6 狭义知识基因与广义知识基因

（1）狭义知识基因。笔者将基于单篇文献的知识基因定义为狭义的知识基因，并在微观层面对文献价值进行结构化提炼与组织。在应用上，单篇文献全文细粒度结构化知识基因提取可提高文献知识发现效率。同时，科技文献知识基因可用于单篇科技文献知识评价，通过后续施引文献的遗传或变异表示论文的内容输出能力。

（2）广义知识基因。上述知识基因的定义与应用基于单篇科技文献，是狭义的知识基因定义。科技文献载体有利于知识基因的量化，当对大量的科技文本进行知识基因遗传与变异计算后，可更为清晰地了解某一学科领域的发展与变化轨迹，出现广义上的知识基因，可用于描述宏观领域的知识进化轨迹。

广义知识基因可以控制知识发展的方向，是某个学科研究方向中文献的共有知识基因体现，如福柯所述："一只看不见的手通过认知模式控制着知识系统"[52]。某些文献在宏观层面存在相似知识基因，并通过知识基因节点紧密连接在一起，形成其研究方向。

7 结语

知识的发展与自然界人类发展存在一定共性。波普尔指出，知识的发展与生物的进化存在着惊人的相似之处，包括地球在内的全部宇宙自然界的第一世界与包括人类所创造的语言、文艺作品、宗教、科学、技术等客观知识世界的第三世界在发展中存在相似规律。基因要素作用于知识组织系统的两大重要作用机制为遗传与变异机制。随着生命科学的飞速发展，人类逐渐掌握基因的结构和作用机理，并发现基因的发展符合自然辩证法，与其他事物发展存在相似规律。基因与知识基因同样符合相同的自然法则规律，通过对知识基因理论的研究，可使我们更加准确把握知识演化规律，促进科学技术的迭代发展。

参考文献：

[1] MARTIN W B. Atoms, bytes and genes public resistance and techno-scientific responses[M]. New York: Routledge,2015:1.
[2] MANDELBROT B. How long is the coast of Britain? statistical self-similarity and fractional dimension [J]. Science, 1967, 156 (3775):636－638.
[3] SIDDHARTHA M. The gene: an intimate history[M] New York: Scribner,2017:121.
[4] 道金斯. 自私的基因[M]. 卢允中,译. 长春:吉林人民出版社,1998:192.
[5] JAMES G. The information: a history, a theory, a flood[M]. New York :Books,2012:307－320.
[6] HERRNSTEIN R, MURRAY C. The bell curve: intelligence and class structure in American life. [J]. Transforming Anthropology, 2010, 6(1/2):87－89.
[7] SEN S K. A note on the idea gene and its relevance to information science[J]. ALIS, 1981, 28(1/4): 97－102.
[8] BLACKMORE S. The meme machine[M]. Oxford: Oxford Paperbacks, 2000.
[9] AUNGER R. The electric meme: a new theory of how we think [M]. New York:Simon and Schuster, 2002.
[10] DALTON C, JASON F. The genome factor: what the social genomics revolution reveals about ourselves, our history, and the future [M]. Princeton: Princeton University Press,2017:4,284.
[11] 李伯文. 论科学的"遗传"和"变异"[J]. 科学学与科学技术管理,1985(10):21－25.
[12] 刘植惠. 知识基因理论新进展[J]. 情报科学,2003(12):1243－1245.
[13] 刘植惠. 知识基因理论的由来、基本内容及发展[J]. 情报理论与实践,1998(2):8－13.
[14] 刘植惠. 知识基因探索(一)[J]. 情报理论与实践,1998(1):63－65.
[15] SUN X L, DING K. Identifying and tracking scientific and technological knowledge memes from citation networks of publications and patents[J]. Scientometrics,2018,116(3): 1735－1748.
[16] 刘则渊. 知识基因论视野下的"新兴研究领域识别计量"著作－《新兴研究领域识别计量》序言[M]. 北京:科学出版社,2017: ⅰ－ⅵ.
[17] 和金生,吕文娟. 知识基因论的源起、内容与发展[J]. 科学学研究,2011,29(10):1454－1459.
[18] 许琦,顾新建. 一种基于 Subject-Action-Object 三元组的知识基因提取方法[J]. 浙江大学学报(工学版),2013,47(3):385－399.
[19] 孙晓玲,丁堃. 基于知识基因发现的科学与技术关系研究[J]. 情报理论与实践, 2017,40(6): 23－26,17.
[20] 逯万辉,谭宗颖. 基于知识基因游离与重组的领域主题演化研究[J]. 情报理论与实践,2019,42(2):101－107.
[21] 吴军. 见识:商业的本质和人生的智慧[M]. 北京:中信出版社,2017:42－46.

[22] 索传军,盖双双.知识元的内涵、结构与描述模型研究[J].中国图书馆学报,2018,44(4):54-72.
[23] RICE W R. Sex chromosomes and the evolution of sexual dimorphism[J]. Evolution, 1984, 38(4):735-742.
[24] 傅荣贤.论古代提要和现代摘要的文献观[J].图书情报工作,2016,60(6):26-31.
[25] 祝清松,冷伏海.引文内容分析方法研究综述[J].情报资料工作,2013(5):39-43.
[26] 徐雷,潘珺.知识网络等相关概念比较分析[J].情报科学,2017,35(12):10-15.
[27] RASKIN R. Enabling semantic interoperability for earth system science[C]// American Geophysical Union. AGU Fall Meeting Abstracts. New York: American Geophysical Union. 2004:11-16
[28] 张运良,徐硕,朱礼军,等.汉语科技词系统——一种可用于科技信息资源深度内容分析的语义资源[J].图书情报工作,2011,55(4):100-105.
[29] MA X, CARRANZA E J M, WU C, et al. A SKOS-based multilingual thesaurus of geological time scale for interoperability of online geological maps[J]. Computers & geosciences, 2011, 37(10):1602-1615.
[30] MOINE M P, VALCKE S, LAWRENCE B N, et al. Development and exploitation of a controlled vocabulary in support of climate modelling[J]. Geoscientific model development, 2014, 7(2):479-493.
[31] SU Y, ANDREWS J, HUANG H, et al. Reengineering of MeSH thesauri for term selection to optimize literature retrieval and knowledge reconstruction in support of stem cell research[J]. BMC medical informatics and decision making, 2016, 16(1):54.
[32] MORAVCSIK M J, MURUGESAN P. Some results on the function and quality of citations[J]. Social studies of science, 1975, 5(1):86-92.
[33] GARZONE M, MERCER R E. Towards an automated citation classifier[J]. Lecture notes in computer science, 2000, 1822:337-346
[34] SPIEGEL-RÖSING I. Science studies: bibliometric and content analysis[J]. Social studies of science, 1977, 7(1):97-113.
[35] OPPENHEIM C, RENN S P. Highly cited old papers and the reasons why they continue to be cited[J]. Journal of the American Society for Information Science, 1978, 29(5):225-231.
[36] RADOULOV R. Exploring automatic citation classification[D]. Waterloo, ON, Canada: University of Waterloo, 2008:33-37.
[37] 陆伟,孟睿,刘兴帮.面向引用关系的引文内容标注框架研究[J].中国图书馆学报,2014,40(6):93-104.
[38] 刘杰,秦春秀,赵捧未,等.基于知识元的科技文本资源内容组织方法[J].情报理论与实践,2018,41(4):128-133.
[39] SHERWIN B. N. How we eie: reflections of life's final vhapter, New Edition[M]. New York: Vintage Books,1995:89.
[40] MONOD J. A biologist's world view(Book Reviews: Chance and Necessity. An essay on the natural philosophy of modern biology)[J]. Science, 1972, 175(4017):49-50.
[41] DENNETT D C. Consciousness explained[M]. New York: Little, Brown and Company, 1991.
[42] VOOS H, DAGAEV K S. Are all citations equal? Or, did we Op. Cit. your idem? [J]. Journal of academic librarianship, 1976, 1.
[43] TEUFEL S, SIDDHARTHAN A, DAN T. Automatic classification of citation function[C]// Proc. 2006 conference on empirical methods in natural language processing. Stroudsburg: Association for Computational Linguistics, 2006:103-110.
[44] LIPETZ B A. Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators[J]. Journal of the Association for Information Science & Technology, 2014, 16(2):81-90.
[45] HERLACH G. Can retrieval of information from citation indexes be simplified? Multiple mention of a reference as a characteristic of the link between cited and citing article[J]. Journal of the Association for Information Science & Technology, 2014, 29(6):308-310.
[46] ZHU X, TURNEY P, LEMIRE D, et al. Measuring academic influence: Not all citations are equal[J]. Journal of the Association for Information Science & Technology, 2015, 66(2):408-427.
[47] 常思敏.科技论文中冗余参考文献分析[J].出版科学,2015,23(1):43-45.
[48] MACROBERTS M H, MACROBERTS B R. Quantitative measures of communication in science: a study of the formal level[J]. Social studies of science, 1986, 16(1):151-172.
[49] CHUBIN D E, MOITRA S D. Content analysis of references: adjunct or alternative to citation counting? [J]. Social studies of science, 1975, 5(4):423-441.
[50] 刘盛博,丁堃,张春博.基于引用内容性质的引文评价研究[J].情报理论与实践,2015,38(3):77-81.
[51] HASSAN S U, SAFDER I, AKRAM A, et al. A novel machine-learning approach to measuring scientific knowledge flows using citation context analysis[J]. Scientometrics, 2018, 116(4):1-24.
[52] 福柯.知识考古学[M].谢强,马月,译.北京:三联书店,1998:202-215.

作者贡献说明:
白如江:负责提出论文框架与设计研究思路;
张庆芝:负责收集资料与撰写论文;
孙一钢:负责研究思路审阅与论文定稿。

白如江，张庆芝，孙一钢.科技文献知识基因表达及遗传与变异研究.[J].图书情报工作,2020,64(4):78-87.

A Study of Knowledge Meme Heredity and Mutation in Academic Paper

Bai Rujiang[1,2] Zhang Qingzhi[3] Sun Yigang[1]
[1] National Library of China, Beijing 100081
[2] Institute of Scientific and Technical Information, Shandong University of Technology, Zibo 25000
[3] Department of Information Management, Peking University, Beijing 100871

Abstract: [**Purpose/significance**] The accumulation and inheritance of knowledge promotes the development of human society. This paper proposes to study the inheritance and variation of knowledge by using the knowledge gene of scientific and technological literature, in order to have a more intuitive and comprehensive perspective on the inheritance and development of knowledge. [**Method/process**] By analyzing the narrow and broad definitions of knowledge genes, the research significance of knowledge genes was determined and their specific research objects were discussed. Two ways of expression of knowledge genes in scientific and technological literature were proposed, and the types of knowledge genes in scientific and technological literature were analyzed. The main factors affecting the inheritance and variation of knowledge genes were summarized, and the inheritance of knowledge genes was designed. And the method of variation. [**Result/conclusion**] The identification of knowledge genes in scientific and technological literature can effectively reveal the knowledge inheritance and iteration between different documents, and promote the development and perfection of the theoretical system of knowledge memes.

Keywords: meme knowledge meme knowledge meme heredity knowledge meme mutation content analysis

《图书情报工作》2020 年选题指南

[**编者按**]本选题指南是根据本刊的定位、性质与发展需要，结合图情档学科前沿热点及当前与未来需要解决的重要问题，邀请本刊编委和青年编委为本刊策划定制，再经编辑部整理、修改和补充而形成的。这是本刊 2020 年度关注、报道的重点领域(包括但不限于这些选题)，供作者选题和研究以及向本刊投稿时的参考和借鉴。

1. 中国特色图情档学科体系、学术体系、话语体系建设
2. 图情档一级学科建设与融合发展战略
3. 图书馆“十四五”规划编制的关键问题
4. 国家文献信息资源保障能力及其建设
5. 开放科学背景下信息资源建设问题
6. 全民阅读中图书馆的定位与担当
7. 图书馆空间服务的理论与实践
8. 嵌入式学科服务的绩效评价与管理
9. 公众科学、科学素养与泛信息素养
10. 图书馆服务本科教育的模式与能力
11. 图书馆文化传承与文化育人的理论与实践
12. 图书馆出版与出版服务
13. 新媒体时代图书馆科学传播的功能与实践
14. 图书馆营销推广的战略与策略研究
15. 图书馆泛合作研究的实践与理论
16. 国家区域发展战略下图书馆联盟建设与创新服务
17. 网络空间治理的情报学问题
18. 知识产权信息服务能力与效果评估
19. 信息分析中的新技术与新方法
20. 情报服务标准化与评价
21. 数字人文与数字学术的研究与实践
22. 人工智能在图情档中的应用
23. 图书馆智能服务与智慧服务
24. 开放数据生态中的元数据发展模式研究
25. 开放科学数据行为及其模型构建
26. 数据资源建设与数据馆员能力建设
27. 大数据时代信息组织与知识组织
28. 科学数据管理与服务
29. 学术成果监测与学科竞争力分析
30. 情报计算(计算情报)的理论与方法
31. 情报分析服务质量与效能评价
32. 情报研究与智库研究的关系
33. 科学与技术前沿分析理论与方法
34. 健康中国 2030 战略下的健康信息学
35. 人机交互行为及服务模式创新
36. 图情档在新型智库建设中的作用机制
37. 智能信息服务的理论和方法
38. 数字公共文化资源、服务与体系建设
39. 数据时代政务信息资源管理和开发利用
40. 数字档案馆生态系统治理策略
41. 档案数据治理理论与治理体系
42. 政府数据开放平台应用与评价
43. 社会记忆视角下档案信息资源整理、保护与开发
44. 民族文献遗产产业化开发与利用
45. 图情档学科教育模式与人才培养能力